Statistical-Dynamical Seasonal Forecast of North Atlantic and U.S. Landfalling Tropical Cyclones using the High-Resolution GFDL FLOR Coupled Model

Hiroyuki Murakami[1, 2], Gabriele Villarini[3],
Gabriel A. Vecchi[1, 2], and Wei Zhang[1, 2]

[1]National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

[2]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA

[3]IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA, USA

Submitted to *the Monthly Weather Review* on 4th September 2015

Corresponding author address: Hiroyuki Murakami, NOAA/GFDL, 201 Forrestal Rd., Princeton, NJ 08540-6649
E-mail: hir.murakami@gmail.com
Tel: 609-452-5824

Abstract

Retrospective seasonal forecasts of North Atlantic tropical cyclone (TC) activity over the period 1980-2014 are conducted using a GFDL high-resolution coupled climate model (FLOR). The focus is on basin-total TC and U.S. landfall frequency. The correlations between observed and model predicted basin-total TC counts range from 0.4 to 0.6 depending on the month of the initial forecast. The correlation values for the U.S. landfalling activity based on individual TCs tracked from the model are smaller and between 0.1 and 0.4. Given the limited skill from the model, statistical methods are used to complement the dynamical seasonal TC prediction from the FLOR model. Observed and predicted TC tracks were classified into four groups using the fuzzy c-mean clustering to evaluate model's predictability in observed classification of TC tracks. Analyses revealed that the FLOR model has the largest skill in predicting TC frequency for the cluster of TC which tracks over the Caribbean Ocean and the Gulf of Mexico.

New hybrid models to improve the prediction of observed basin-total TC and landfall TC frequencies are developed. These models use large-scale climate predictors from the FLOR model as predictors for generalized linear models. The hybrid models show considerable improvements in the skill in predicting the basin-total TC frequencies relative to the dynamical model. The new hybrid model shows correlation coefficients as high as 0.75 for basin-wide TC counts from the first two lead months and retains values around 0.52 even at the 6-month lead forecast. The hybrid model also shows comparable or higher skill in forecasting U.S. landfalling TCs relative to the dynamical predictions. The correlation coefficient is about 0.5 for the 2-6 month lead times.

1. Introduction

Tropical cyclones (TCs) were the most costly natural disaster to affect the United States (U.S.) over the period 1980–2011 (Pielke et al. 2008; Smith and Katz 2013). According to Smith and Katz (2013), TCs were responsible for over $400 billion in damages over the period (Consumer Price Index), which correspond to 47.4% of all the damage caused by all the natural disasters responsible for $1+ billion combined. Smith and Katz (2013) also reported apparent increasing trends in both the annual frequency of billion- dollar events and in the annual aggregate loss from these events. Therefore, predicting TC activity at seasonal time scales is a topic of large scientific and socio-economic interest.

Since Gray (1984a, b) first attempted seasonal forecasts of TC activity for the North Atlantic (NA), tremendous effort has been devoted to construct and improve statistical models in which observed large-scale climate indices ahead of the hurricane season are used to predict subsequent summertime basin total TC frequency (Gray et al. 1992, 1993, 1994; Klotzbach and Gray 2004, 2009; Elsner and Jagger 2006; Klotzbach 2008) and landfalling TCs (Lehmiller et al. 1997; Klotzbach and Gray 2003, 2004; Saunders and Lea 2005; Elsner et al. 2006; Klotzbach 2008; Jagger and Elsner 2010). However, most of the current statistical seasonal forecasts show skill for forecasts starting from April and later for the subsequent TC season in July-November (e.g., Elsner and Jagger 2006), and prediction skill is limited when the lead time increases and the target region is smaller than the entire North Atlantic (Lehmiller et al. 1997; Klotzbach and Gray 2012).

Recent advances in dynamical modeling and computational resources have enabled prediction using high-resolution dynamical models [see review in Camargo et al. (2007)]. These models showed significant skill in predicting basin-total TCs for seasonal prediction

(e.g., Vitart and Stockdale 2001; Vitart 2006; Vitart et al. 2007; LaRow et al. 2008; Camargo and Barnston 2009; LaRow et al. 2010; Zhao et al. 2010; Alessandri et al. 2011; Chen and Lin 2011, 2013; Vecchi et al. 2014; Camp et al. 2015), with correlation values up to 0.96 between observed and the predicted North Atlantic TC counts over the 2000–2010 period (Chen and Lin 2011, 2013). However, predicting U.S. landfall frequency using dynamical models remains challenging even though it is of paramount societal and scientific importance (Vecchi and Villarini 2014). Vecchi et al. (2014) and Camp et al. (2015) found some predictive skill for TC landfall in the Caribbean, but limited skill for U.S. landfall frequency.

Some of the limitations of dynamically forecasting TCs can be alleviated using so-called "hybrid predictions" or "statistical-dynamical predictions." In the hybrid predictions, a statistical model is constructed using the empirical relationship between observed TC activity and predicted large-scale parameters simulated by a dynamical model. Using the statistical model, future TC activity is then predicted given the large-scale parameters predicted by a dynamical model (e.g., Zhao et al. 2010; Wang et al. 2010; Vecchi et al. 2011, 2013, 2014). For example, previous studies showed that basin total North Atlantic TC activity substantially correlated with relative sea surface temperature (SST) anomalies (i.e., local SST anomaly relative to tropical mean anomaly) in observations (e.g., Latif et al. 2007; Swanson 2008; Vecchi et al. 2008; Villarini et al. 2010, Villarini and Vecchi 2012) and dynamical models (Zhao et al. 2010; Villarini and et al. 2011; Murakami et al. 2012; Knutson et al. 2013; Ramsay and Sobel 2011). Using these simulated/predicted SST anomalies as predictors, previous studies achieved substantial skill in predicting basin-total TC frequency (Zhao et al. 2010; Vecchi et al. 2011, 2013, 2014), basin-total power dissipation index (PDI) and accumulated cyclone energy (ACE)(Villarini and Vecchi 2013) compared to dynamical

models. However, while these studies showed that it is possible to skillfully forecast North Atlantic TC activity at the basin scale, little is known about the applicability of hybrid systems at much more regional scales.

Vecchi et al. (2014) reported that the high-resolution dynamical model that will be used in this study has higher skill in predicting TCs near the coastline of the Gulf of Mexico and Caribbean Sea relative to those near the coastline of northeastern United States (see their figure 13). This indicates that models may have higher skill in predicting/simulating one or more groups of TC tracks. If any hybrid models could improve predictions for the groups with poor forecasting skill, we could improve prediction skill for landfall TC frequency as well as basin total TC frequency. Moreover, finding predictors in the way of constructing hybrid model will help the understanding of the potential physical mechanisms responsible for U.S. landfalling TCs. Kossin et al. (2010) classified all NA TC tracks into four clusters, revealing distinct characteristics for each cluster in terms of their tracks and genesis locations, seasonality, and relationship between frequency of TCs and climate variability. Colbert and Soden (2012) classified TC tracks into three groups (straight moving, recurving landfall, or recurving ocean) highlighting differences in the climate conditions associated with each one of them. However, there is no information about how predictable these TC clusters are.

In this study, we first examine the predictability of observed basin-total TC frequency of observed TC clusters. Second, we attempt to construct a hybrid model to improve the prediction skill in TC frequency for each cluster, which in turn leads to the improvements in predicting basin-total TC frequency. Third, we examine observed and predicted TC landfall ratio, and construct a hybrid model to improve prediction skill in TC landfall ratio. Finally, we

show the forecasting skill in landfall TC frequency over the United States predicted by FLOR and compare the results from our newly developed hybrid model.

The remainder of this paper is organized as follows. Section 2 describes the models and data used in this study. Section 3 assesses the performance of the hybrid models in predicting TCs within each considered TC cluster, TC landfall ratio, and TC landfall frequency over the United States and compares these results to the dynamical model. Finally, Section 6 provides a summary of the results.

2. Methods

Throughout this study, we focus on the prediction of North Atlantic TCs during July–November because about 84% of all storms occurred during these months over the 1980–2014 period. We focus on tropical storms or more intense cyclones (wind speed >34 kt), and these storms are defined as TCs. The targeted prediction is the frequency of basin-total TCs and landfalling TCs along the U.S. coastline. In this section, dynamical models, observed data, and the TC detection algorithms are described.

a. Dynamical model

The dynamical model used for retrospective seasonal forecasts is the Forecast-oriented Low Ocean Resolution (FLOR; Vecchi et al. 2014; Jia et al. 2015a) of the Geophysical Fluid Dynamics Laboratory (GFDL) Coupled Model version 2.5 (CM2.5; Delworth et al. 2012). FLOR comprises 50-km mesh atmosphere and land components, and 100-km mesh sea ice and ocean components. For each year and each month in the period 1980–2014, 12-month duration predictions were performed after initializing the model to observationally constrained

conditions. Here, we defined forecasts from July, June,…, January, December initial conditions as the lead-month (L) 0, 1, …, 6, 7 forecasts for the predictions of TC activity in subsequent summer (July–November).

The 12-member initial conditions for ocean and sea ice components were built through a coupled ensemble Kalman filter (EnKF; Zhang and Rosati 2010) data assimilation system developed for the GFDL Coupled Model version 2.1 (CM2.1; Delworth et al. 2006; Wittenberg et al. 2006; Gnanadesikan et al. 2006), whereas those for atmosphere and land components were built from a suite of SST-forced atmosphere-land-only simulations to the observed values using the components in FLOR. Therefore, the predictability comes entirely from the ocean and sea ice, and may be thought of as a lower bound on the potential prediction skill of a model because predictability could also arise from atmospheric (particularly stratospheric) and land initialization. During the simulation using FLOR, simulated temperature and wind stress are adjusted using so-called "flux-adjustment" in which model's momentum, enthalpy and freshwater fluxes from atmosphere to ocean are adjusted to bring the model's long-term climatology of SST and surface wind stress closer to observations and improve simulations of TCs and precipitation (Vecchi et al. 2014; Delworth et al. 2015).

b. Observational datasets and detection algorithm for tropical cyclones

The observed TC "best-track" data were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010) and used to evaluate the TC simulations in the retrospective seasonal predictions. We also use the UK Met Office Hadley Centre SST product (HadISST1.1; Rayner et al. 2003) as observed SST.

136 Model-generated TCs were detected directly from 6-hourly output using the following
137 tracking scheme documented in Murakami et al. (2015). In the detection scheme, the flood fill
138 algorithm is applied to find closed contours of some specified negative sea level pressure
139 (SLP) anomaly with warm core (1K temperature anomaly). The detection scheme also
140 considers satisfaction of duration of 36 consecutive hours for which TC candidate should
141 maintain warm core and wind criteria (15.75 m s^{-1}).

142

143 3. Results

144 *a. Clustering TC tracks and forecasting skill by a dynamical model*

145 We first applied a clustering algorithm to observed TC tracks (Fig. 1, green tracks). Fig.1
146 The cluster technique used here is the fuzzy c-means clustering developed by Kim et al.
147 (2011). Fuzzy clustering has been known to produce more natural classification results for
148 datasets such as TC tracks that are too complex to determine their boundaries of distinctive
149 pattern (Kim et al. 2011). Following Kossin et al. (2010), the final number of clusters is equal
150 to 4, yielding early recurving TCs (CL1), the Gulf of Mexico and Caribbean TCs (CL2),
151 subtropical (or extratropical transition)-type TCs (CL3), and classic "Cape Verde hurricanes"
152 (CL4). Each cluster receives a comparable number of the total storms as shown in the
153 fractional ratio between 20% and 28%. When compared to Kossin et al. (2010), CL2 and CL3
154 are similar to the two clusters in their study, whereas CL1 and CL4 are different from their
155 results. This could be due to the different study period. Kossin et al. (2010) used observed data
156 for the period 1950–2007, whereas we focus on the period 1980–2014. When we extend the
157 fuzzy clustering analysis to 1958–2014, we still obtain the same clustering groups as shown in

158 Fig. 1, indicating that the differences from Kossin et al. (2010) may be due to the difference in
159 clustering methodology, rather than differences in study period.

160 Second, we assigned predicted TCs to one of the observed TC clusters. Regardless of
161 any lead-month forecasts and ensemble members, we computed the root-mean-square error
162 (RMSE) between the predicted (black track) and observed mean (red track) TC track for each
163 TC cluster. To compute RMSE, we interpolate every TC track into 20 segments with equal
164 length following Kim et al. (2011). We assign the predicted TC track to the TC cluster with
165 the minimum RMSE. An alternative way is to conduct the cluster analysis using the combined
166 data of observed and predicted TC tracks. However, because we obtained similar results to the
167 method above (figure not shown), we will use the RMSE for the assignment. The results for
168 assigned TC tracks are shown in Fig. 1 as black tracks. Although the dynamical model
169 predicts fractional ratios of TC frequency for CL1 and CL3 similar to the observations (about
170 20%), it slightly overestimates (underestimates) the fractional ratio for CL2 (CL4). Figure 2 Fig.2
171 shows forecast skill in predicting TC frequency for each cluster and for each lead month by
172 the dynamical model in terms of rank correlation (Fig. 2a) and RMSE (Fig. 2b). For the
173 sample size of 35 years (i.e., 1980–2014), correlations of 0.33 and 0.43 are statistically
174 significant at the 5% and 1% levels. Rank correlation for the basin total TC counts (black line
175 in Fig. 2a) is about 0.6 for lead time L=0–2, and decreases to about 0.4 for L=5–7. Vecchi et al.
176 (2014) also reported similar results for the correlations for the basin-total frequency. RMSE
177 for the basin total TC counts (black line in Fig. 2b) is about 5–7. This large RMSE is mainly
178 because of the underestimation in predicting TC frequency as also reported in Murakami et al.
179 (2015). Shorter lead-month predictions show larger RMSE (black line in Fig. 2b). Although
180 further investigation is required, this may be related to initial spin-ups due to the cold SST bias

181 in the initial condition in the tropical North Atlantic, which is inherited from CM2.1 for the
182 initialization. During the predictions for the first few months, FLOR tries to adjust the cold
183 bias through the flux adjustments; however, it may take a few months to adjust the cold biases.
184 Among the four TC clusters, the dynamical model shows relatively higher skill in
185 predicting CL2, followed by CL1 and CL4. The higher skill in predicting CL2 is consistent
186 with Vecchi et al. (2014), who reported that FLOR has higher skill in predicting TCs near the
187 coastline of the Gulf of Mexico and Caribbean Sea. On the other hand, the FLOR predictions
188 for CL3 show the lowest skill, indicating that the prediction of TCs that undergo extratropical
189 transition remains challenging for dynamical models [see also Jones et al. (2003)]; it is unclear
190 whether this reflects a deficiency in the models and initialization, or an inherent limit to the
191 predictability of the year-to-year variations of CL3 storms.
192
193 *b. Correlations between observed TC frequencies and predicted large-scale parameters*
194 Section 3a showed that the dynamical model has the lowest skill in predicting CL3 and
195 CL4 TCs. If these biases could be improved, prediction of basin-total frequency and
196 landfalling TC frequency could potentially be improved as well. For this purpose, we start by
197 constructing a hybrid model in which observed TC frequency is regressed and predicted using
198 some key large-scale parameters simulated by the dynamical model for each cluster. To
199 identify the key parameters, we first investigate correlations between observed TC frequency
200 and large-scale parameters. The large-scale parameters considered are relative SST (*RSST*),
201 geopotential height at 500 hPa (Z_{500}), and zonal component of vertical wind shear (200–850
202 hPa, *WS*). The *RSST* is defined as the local SST anomaly subtracted from the tropical mean
203 (30°S–30°N) SST anomaly. We have performed a preliminary investigation of including other

parameters such as mid-level relative humidity, low-level relative vorticity, SLP, steering flow among others; however, these parameters did not show significant correlations with TC frequency for any clusters (figure not shown). Also, we prefer parsimonious models that incorporate a smaller number of predictors in order to help avoid an over-fitting problem when a hybrid model is constructed.

Figure 3 shows correlation map between the time-series of observed TC frequency for each cluster and the three large-scale parameters computed for each 1° × 1° grid box within the global domain during the peak season. For *RSST* (Fig. 3a–d), all clusters except for CL3 show higher positive correlations in the tropical NA, which is consistent with previous studies (e.g., Villarini et al. 2010; Vecchi et al. 2011; Villarini and Vecchi 2012). CL2 and CL4 also show the La Niña-like pattern in the Pacific, indicating that TC frequency for these clusters increase during La Niña years. Although CL1 does not show the La Niña-like pattern clearly, the cluster shows negative correlation in the subtropical Pacific. CL3 is unique with respect to the other clusters because there is no significant pattern in the correlation even in the tropical North Atlantic, indicating that CL3 is insensitive to local SST anomaly. As for Z_{500} (Fig. 3e–h), there are higher positive correlations in the subtropical central pacific for CL1, CL2, and CL4. A preliminary investigation implies that this correlation is related to the Pacific/North American (PNA) pattern. We found that when the anomaly of Z_{500} is positive in the box of Fig. 3e, f, and h, Z_{500} in the subtropical North Atlantic (30–50°N, 55–75°W) is negative through a series of wave train along the subtropical westerly jet. Also, during the positive phase, convection is active in the tropical North Atlantic and western African coast, leading to more frequent easterly waves and TC development associated with the enhanced convection. On the other hand, CL3 shows no correlation with Z_{500}, which is again largely different from other

Fig.3

227 clusters. For the *WS* (Fig. 3i–l), there are higher negative correlations in the tropical NA for
228 CL1, CL2, and CL4, which is reasonable because TC activity is unfavorable under strong
229 vertical wind shear. However, it is intriguing that CL3 is not sensitive to vertical wind shear.

230 CL3 does not show significant correlations with any parameters, suggesting that CL3
231 may have a substantially larger stochastic element to its variability than the other clusters, and
232 may thus be inherently less predictable. However, we want to identify any key large-scale
233 parameters to construct a hybrid model. When correlation is computed between observed TC
234 frequency and *RSST*, observations (Fig. 4b) and dynamical model (Fig. 4a) show relatively Fig.4
235 higher correlations in the four domains. Although physical mechanisms explaining the
236 relationship between Atlantic CL3 TCs and the remote SSTs are difficult to interpret, we
237 utilized *RSST* in these four domains as predictor for CL3 TCs.

238 The domains of the predictors used for the hybrid model are shown in the rectangles in
239 Fig. 3. The dynamical model should also have significant forecast skill in predicting these
240 large-scale parameters for each domain. Figure 5 shows anomaly correlations for the lead- Fig.5
241 months 0, 3, and 6, respectively, for each parameter. The red shaded area is the region where
242 anomaly correlation exceeds 0.5, revealing that the dynamical model has skill in predicting the
243 large-scale parameters for each domain used for predictors, even for the lead-month 6. The
244 skill in predictions and the correlation with respect to the observations justifies the use of the
245 large-scale parameters in the domains as the predictors for the hybrid model.

246

247 *c. Hybrid Poisson regression model*

248 Using the predictors discussed in Section 3b, a Poisson regression model (e.g.,
249 Villarini et al. 2010; Elsner and Jaeger 2013) is constructed to predict observed TC frequency

for each cluster using large-scale parameters predicted by the dynamical model of FLOR. First of all, the probability of TC frequency x is obtained when the mean frequency (i.e., rate) λ is given as follows.

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \text{ where } x = 0, 1, 2, \ldots \text{ and } \lambda > 0. \quad (1)$$

The Poisson regression model is expressed as

$$\log(\lambda) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \quad (2)$$

There are p predictors and $p+1$ parameters and a logarithmic link function. First we determine β_p given the observed λ and simulated x_p (regression or training). Then, the cross validation is performed to evaluate the model skill. Here we apply so-called "leave-one-out cross validation" (LOOCV; Elsner and Jaeger 2013). In the LOOCV, we first exclude a single year of observations and predictors; then, we determine the coefficients of the Poisson regression model using remaining years. Using the model, TC count for the excluded year is predicted. This is done for 35 years, removing each year's data point successively.

Figure 6 reveals results of training (Fig. 6a–d) and LOOCV (Fig. 6e–h) for each cluster at lead-month 0. To compare skill in these hybrid models with the dynamical model, Fig. 7 shows comparisons of rank correlations (Fig. 7a,c) and RMSE (Fig. 7b,d) between the dynamical model (solid lines) and LOOCV (dashed lines). Predictions for all of clusters using the hybrid approach are improved in LOOCV in terms of RMSE for every lead month. Although CL1 was not improved in terms of correlation, most of the clusters show improvements in simulating observed interannual variation. When these predicted TC frequencies are summed up, we derive the basin-total TC frequency. The basin total frequency also shows higher skill in the hybrid than the dynamical model (Fig. 7c,d). We obtain a maximum correlation coefficient of 0.76 at lead-month 1 and the minimum correlation is still

Fig.6

Fig.7

273 high at 0.54 at lead-month 7. On the other hand, the values of the correlation coefficient for
274 dynamical model are 0.63 and 0.35, respectively, highlighting the improvements introduced
275 by the hybrid model in predicting basin-total TC frequency.
276
277 *d. Correlations between observed TC landfall ratio and predicted large-scale parameters*
278 Section 3c revealed that the skill in predicting basin-total TC frequency is higher in the
279 hybrid than in the dynamical model. Therefore, skilful forecasting of the fraction of total TCs
280 making landfall in the United States could lead to accurate predictions of TC landfall activity
281 when combined with prediction of basin-total TC frequency. In this section, we first
282 investigate the physical drivers for the observed landfall ratio. The landfall domain defined in
283 this study is the coastal region of the United States as identified in the blue region in Fig. 1. In
284 this study, once a TC propagates in the blue region in Fig. 1, we count one for TC landfall
285 frequency regardless of multiple landfall events for the same TC. Figure 8 shows the Fig.8
286 interannual variation of basin-total TC frequency (red), landfall TC frequency in the United
287 States (blue), landfall ratio (black), and Niño-3.4 index (green) in the observations. The Niño-
288 3.4 index is obtained from the mean SST anomaly in the region bounded by 5°N and 5°S, and
289 between 170°W to 120°W. The rank correlation between basin-total TC frequency and
290 landfall ratio is 0.08, indicating that there is no strong linear relationship between the two
291 variables. Indeed, while there were 18 TCs in 2010, which was the third largest TC frequency
292 during the period 1980–2014, only one of them made landfall in the United States that year.
293 The observed relationship between TC landfall ratio and climate indices was analysed
294 by Villarini et al. (2012). They constructed a statistical model to predict landfall ratio using
295 three predictors [May–June mean North Atlantic Oscillation (NAO), the Southern Oscillation

296 index (SOI), and tropical mean SST (30S°–30°N)]. Here we computed the correlation
297 coefficient between observed TC landfall ratio and simulated large-scale parameters of SST
298 and May-June mean SLP as an indication of effect of NAO. However, we could not find any
299 significant correlation using SLP (figure not shown). Nevertheless, we found a La Niña-like
300 pattern in the correlation for SST (Fig. 9), indicating that TC landfall ratio tends to be higher
301 (lower) during La Niña (El Niño) years. The increased landfall ratio during La Niña years is
302 consistent with Bove et al. (2008) who examined the effects of El Niño on U.S. landfalling
303 hurricanes and found that the probability of U.S. hurricanes increased from 28% during El
304 Niño years to 66% during La Niña years. However, although the rank correlation between
305 landfall ratio (black line in Fig. 8) and Niño-3.4 index (green line in Fig.8) is negative (i.e., –
306 0.24), the correlation is not statistically significant at 90% level, indicating that landfall ratio is
307 only slightly correlated with the La Niña conditions.
308
309 *e. Hybrid binomial regression model*
310 As shown in section 3d, SST anomalies in the tropical Pacific are correlated with TC
311 landfall ratio for the United States as an indication of La Niña years. Although the correlation
312 is not high, we use the SST anomalies in the domain shown Fig. 9 as a predictor for the hybrid Fig.9
313 model for predicting TC landfall ratio using a binomial regression model (Villarini et al. 2012).
314 Following Villarini et al. (2012), let us define Y_1 and Y_2 as two Poisson random variables with
315 means of μ_1 and μ_2. Let us define m as their sum ($m = Y_1 + Y_2$), which also follows a Poisson
316 distribution with mean equal to $\mu_1 + \mu_2$. In the case of this study, m represents the basin-total
317 TC frequency, whereas Y_1 represents the frequency of landfall TCs over U.S. Given m, the
318 distribution of Y_1 can be written as

$$f(Y_1 = y|\mu) = \frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)}\mu^y(1-\mu)^{(m-y)}, \quad (3)$$

where $\mu = \mu_1/(\mu_1 + \mu_2)$. The mean and the variance of Y_1/m are μ and $\mu(1-\mu)$, respectively. Similar to what is described in Eq. (2), we can relate the parameter μ to a vector of p predictors:

$$\log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p. \quad (4)$$

The dependence of μ on the predictors can be written as

$$\mu = \frac{\exp(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}{1+\exp(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}. \quad (5)$$

Here we consider one predictor (i.e., p=1) of SST as discussed above. Similar to the procedure described in Section 3c, we first determine β_p given the observed μ and simulated x_p. Then, the LOOCV is performed to evaluate the hybrid model.

Figure 10 reveals results of training (Fig. 10a–c) and LOOCV (Fig. 10d–e) for lead-month 0, 3, and 6, respectively. The overall correlation is relatively low (i.e., at most 0.37), indicating that landfall ratio remains difficult to improve even using the hybrid model. To compare skill in the hybrid model with the dynamical model, Fig. 11 shows comparisons of rank correlations (Fig. 11a) and RMSE (Fig. 11b) between the dynamical model (solid line) and LOOCV (dashed line). Although rank correlation is lower when compared with the basin-total TC frequency (Fig. 7c), the hybrid model shows higher skill in predicting landfall ratio than the dynamical mode does. RMSE (Fig. 11b) looks similar between the hybrid and dynamical models, although the hybrid model shows slightly lower RMSE than the dynamical model.

Fig.10

Fig.11

f. Synthesized hybrid model for predicting landfall TCs over the United States

Here we have two hybrid models: the Poisson regression model to predict TC frequency for each cluster, yielding basin total TC frequency by summing all TC clusters (Section 3c); the binomial regression model to predict TC landfall ratio over the United States (Section 3e). By combining the two hybrid models, we can make predictions of TC landfall frequency over United States. A schematic diagram is shown in Fig. 12 for the synthesized hybrid model. Given the key large-scale parameters for a specific year (Step 1), we predict mean TC frequency for each cluster ($\lambda_{1,2,3,4}$) (Step 2). Given the predicted mean λ, random resampling of λ is performed for k times based on the Poisson distribution as shown in Eq. (1), thereby yielding k samples of λ for each cluster (Step 3). For each iteration, λ values for all the clusters are summed up, providing a sample of basin total TC frequency (N) (Step 3). A similar resampling procedure is performed for the TC landfall ratio, yielding k samples of μ (Steps 4–6). For each sample, landfall TC frequency over the United States (X) is computed by multiplying N and μ (Step 7). Based on the k samples for X, we can compute a probabilistic range (e.g., range of 10% bottom bound or 90% top bound) of predicted TC landfall frequency as well as mean X value for each year. Fig.12

Figure 13 shows comparisons between the dynamic model and the synthesized hybrid model in terms of landfall TC frequency over the United States. First of all, the dynamical model systematically underestimates landfall TC frequency, whereas this underestimation is improved in the hybrid model. Moreover, the amplitude of interannual variation is much larger in the hybrid than the dynamical model. For example, the anomalous year of 1998 is well predicted by the hybrid model. On the other hand, the hybrid model significantly overestimates TC landfall frequency in 2010. This year is characterized by La Niña conditions. From Fig. 3a-d and Fig. 9, both TC frequency for each cluster and TC landfall ratio are Fig.13

364 expected to be large for this year. This year is also characterized by negative NAO. Villarini et
365 al. (2012) reported a larger fraction of storms making U.S. landfall during negative NAO
366 phase based on observations, indicating that even if we incorporated NAO index into the
367 hybrid model, the hybrid model would still overestimate 2010 landfall TCs.

368 Figure 14 summarizes the comparisons of rank correlations and RMSE for each lead Fig.14
369 month between the dynamical and hybrid models. The hybrid model shows comparable or
370 higher skill for most of the lead months in terms of rank correlations relative to dynamical
371 forecasts, although correlations show no significant differences in the first three and the last
372 two lead months (Fig. 14a). The hybrid model shows smaller RMSE for all lead months
373 relative to the dynamical forecasts. We can conclude that our new hybrid model retains
374 forecast skill up to lead month 5 with correlation coefficient 0.5 and forecast root mean square
375 error of 2.0 storms per year for U.S. landfalling TCs.

376 We also preliminarily checked the performance of an alternative statistical method in
377 which TC landfall frequency is computed using the constant climatological mean landfall ratio
378 based on observations along with the TC frequency predicted from the Poisson regression
379 model. Although the method shows some improvements in terms of RMSE for the lead month
380 0 and 1 predictions relative to the synthesized hybrid model, the scheme does not show
381 improvements in terms of rank correlation. Accurate predictions for landfall TC frequency
382 seem to be critically dependent on the accurate prediction for landfall ratio in which the
383 present study shows limited skill (Fig. 11).

384

385 # 4. Summary

In this study, we evaluated retrospective seasonal forecasts based on a GFDL high-resolution coupled climate model (FLOR), and we constructed new hybrid models to improve the forecast skill in predicting the frequency of basin-total and U.S. landfalling TCs. First, we classified observed TCs into four groups of TCs using the fuzzy c-means clustering algorithm. Predicted TCs by FLOR are assigned to one of the observed clusters. We found that FLOR has high skill in predicting the Gulf of Mexico and Caribbean TCs (CL2), whereas it has low skill in predicting the subtropical TCs (CL3). The CL3 storms also exhibited limited statistical relationships to large-scale climate conditions, suggesting that the limited prediction skill may reflect limited underlying predictability.

Second, we constructed a hybrid model to predict TC frequency for each cluster using the empirical relationship between observed TC frequency and predicted large-scale parameters by a dynamical model. The hybrid model shows equivalent or higher skill in predicting TC frequency for each cluster relative to the dynamical model. The improvements for each cluster result in improvement in predicting basin-total TC frequency. We obtained maximum and minimum values of the correlation coefficient equal to 0.75 and 0.52 at lead-month 1 and 7, whereas those for counting TCs directly from the dynamical model are 0.63 and 0.35, respectively.

Third, we evaluated retrospective prediction skill for the TC landfall ratio over the U.S., revealing that the dynamical predictions have no skill in predicting the landfall ratio when looking at simulated storms directly. Meanwhile, the observed TC landfall ratio is analyzed, revealing that the landfall ratio has no correlation with basin-total TC frequency. However, the observed interannual variation in landfall TC ratio has a moderate correlation with SST anomaly in the tropical Pacific. This is associated with La Niña-like pattern, indicating that TC

409 landfall ratio is higher during La Niña years. A binomial hybrid model was constructed for
410 better prediction of U.S. landfall TC ratios using the simulated SST anomaly in the tropical
411 Pacific. The hybrid model predicts interannual variations of TC landfall ratio better than the
412 dynamical model does.

413 By combining the two hybrid Poisson and binomial models, the frequency of TC
414 landfall is predicted. The synthesized hybrid model shows comparable or better prediction of
415 TC landfall frequency relative to the dynamical model in terms of RMSE relative to observed
416 TC landfall frequency, correlation between predicted and observed TC landfall frequency, and
417 amplitude of interannual variation. We can conclude that the new hybrid model retains
418 forecast skill up to lead month 5 with correlation coefficient 0.5 and forecast error of 2.0 for
419 TC landfall for U.S.

420 In this study, we used the results of retrospective forecasts by FLOR in which the
421 initial state of the atmosphere and land components are not constrained by observations, while
422 the oceanic component is. We hypothesize that if we initialized the atmospheric component,
423 we might obtain better skill in predicting TC activity in the North Atlantic. Similar tests with
424 FLOR show improved seasonal predictions of land surface conditions with atmospheric
425 initializations (Jia et al. 2015b). Recent studies show that the dynamical models have longer
426 prediction skill of 2-weeks or more due to accurate simulation of intraseasonal oscillations
427 such as the Madden Julian Oscillation (MJO) (e.g., Xiang et al. 2015a,b; Nakano et al. 2015).
428 Further, Murakami et al. (2015) showed that FLOR has the capability of simulating a strong
429 MJO signal. If atmospheric initial conditions contain observed MJO phase and amplitude, the
430 dynamical model may predict TC activity well at least for the shortest lead-month forecast of
431 L=0. Moreover, Murakami et al. (2015) showed better prediction of 1997/1998 TC activity

using the higher resolution version of FLOR (i.e., HiFLOR) in addition to the better simulations of large-scale parameters than FLOR. In the future, we plan to construct a hybrid model using large-scale parameters simulated using HiFLOR to improve the skill in predicting TCs in the North Atlantic.

Acknowledgments This material is based in part upon work supported by the National Science Foundation under Grant AGS-1262099, the Award number NA14OAR4830101 to the Trustees of Princeton University (Gabriele Villarini). This report was prepared by Hiroyuki Murakami under award NA14OAR4830101 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Gabriele Villarini also acknowledges financial support from the USACE Institute for Water Resources. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce. The authors thank Dr. Jan-Huey Chen and Dr. Lakshmi Krishnamurthy for their suggestions and comments.

References

Alessandri, A., A. Borrelli, S. Gualdi, E. Scoccimarro, and S. Masina, 2011: Tropical cyclone count forecasting using a dynamical seasonal prediction system: Sensitivity to improved ocean initialization. *J. Climate,* **24,** 2963–2982.

Bove, M. C., J. B. Elsner, C. W. Landsea, X. Niu, and J.J. O'Brien, 1998: Effect of El Niño on U.S. landfalling hurricanes, revisited. *Bull. Amer. Meteor. Soc.*, **79,** 2477–2482.

Camargo, S. J., A. G. Barnston, P. J. Klotzbach, and C. W. Landsea, 2007: Seasonal tropical cyclone forecasts. *WMO Bull.*, **56,** 297–309.

Chen, J.-H. and S.-J. Lin, 2011: The remarkable predictability of inter-annual variability of Atlantic hurricanes during the past decade. *Geophys. Res. Lett.,* **38,** L11804.

Chen, J.-H. and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate,* **26,** 380–398.

Colbert, A. J., and B. J. Soden, 2012: Climatological variations in North Atlantic tropical cyclone tracks. *J. Climate,* **25,** 657–673.

Delworth, T. L., and co-authors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19,** 643–674.

Delworth, T. L., and co-authors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25,** 2755–2781.

Delworth, T. L., F. Zeng, A. Rosati, G.A. Vecchi, and A. T. Witternberg, 2015: A link betweenthe hiatus in global warming and North American drought. *J. Climate,* **28,** 3834–3845.

Elsner, J. B., J. Murnane, and T. H. Jagger, 2006: Forecasting U.S. hurricanes 6 months in advance. *Geophys. Res. Lett.,* **33,** L10704.

Elsner, J. B., and T. H. Jagger, 2006: Prediction models for annual U.S. hurricane counts. *J. Climate*, **19,** 2935–2952.

Elsner, J. B., and T. H. Jagger, 2013: Frequency models. In "Hurricane Climatology. A modern statistical guide using R." Oxford University Press, pp 161–193.

Gnanadesikan, A., and co-authors, 2006: GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation. *J. Climate*, **19,** 675–697.

Gray W. M., 1984a: Atlantic seasonal hurricane frequency. Part I: El Nino and 30 mb quasi-binnial oscillation influences. *Mon. Wea. Rev.*, **112,** 1649–1668.

Gray W. M., 1984b: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112,** 1669–1683.

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Wea. Forecasting.* **7,** 440–455.

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry, 1993: Predicting Atlantic seasonal hurricane activity by 1 August. *Wea. Forecasting.* **8,** 73–86.

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry, 1994: Predicting Atlantic seasonal hurricane activity by 1 June. *Wea. Forecasting.* **9,** 103–115.

Jones, S. C., and coauthors, 2003: The extratropical transition of tropical cyclones: Forecast challenges, current understanding, and future directions. *Wea. Forecasting.* **18,** 1052–1092.

Jia, L., and co-authors, 2015a: Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. *J. Climate*, **28,** 2044–2062.

Jia, L., G. A. Vecchi, X. Yang, R. Gudgel, T. L. Delworth, W. Stern, K, Paffendorf, S. Underwood, and F. Zeng, 2015b: The roles of radiative forcing, sea surface temperatures,

and atmospheric and land initial conditions in U.S. summer warming episodes. *J. Climate*, submitted.

Lehmiller, G. S., T. B. Kimberlain, and J. B. Elsner, 1997: Seasonal prediction models for North Atlantic basin hurricane location. *Mon. Wea. Rev.*, **125,** 1780–1791.

Kim, H.-S., C.-H. Ho, and P.-S. Chu, 2011: Pattern classification of typhoon tracks using the fuzzy c-means clustering method. *J. Climate*, **24,** 488–508.

Klotzbach, P. J., and W. M. Gray, 2003: Forecasting September Atlantic basin tropical cyclone activity. *Wea. Forecasting.* **18,** 1109–1128.

Klotzbach, P. J., and W. M. Gray, 2004: Updated 6–11-month prediction of Atlantic basin seasonal hurricane activity. *Wea. Forecasting,* **19,** 917–934.

Klotzbach, P. J., and W. M. Gray, 2009: Twenty-five years of Atlantic basin seasonal hurricane forecasts (1984–2008). *Geophys. Res. Lett.,* **36,** L09711.

Klotzbach, P. J., and W. M. Gray, 2012: Qualitative discussion of Atlantic basin seasonal hurricane activity for 2012 [Online available at http://hurricane.atmos.colostate.edu/forecasts/2011/dec2011/dec2011.pdf].

Klotzbach. P. J., 2008: Refinements to Atlantic basin seasonal hurricane prediction from 1 December. *J. Geophys. Res*., **113**, D17109.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neuman, 2010: The international best track archive for climate stewardship (IBTrACS): Unifying tropical cyclone best track data, *Bull. Amer. Meteor. Soc.*, **91**, 363–376.

Knutson, T. R., and coauthors, 2013: Dynamical downscaling projections of twenty-first-century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios. *J. Climate*, **26,** 6591–6617.

LaRow, T. E., Y. –K. Lim, D. W. Shin, E. P. Chassgnet, and S. Cocke, 2008: Atlantic basin seasonal hurricane simulations. *J. Climate*, **21,** 3191–3206.

LaRow, T. E., L. Stefanova, D. –W. Shin, and S. Cocke, 2010: Seasonal Atlantic tropical cyclone hindcasting/forecasting using two sea surface temperature datasets. *Geophys. Res. Lett.,* **37,** L02804.

Latif, M., N. Keenlyside, and J. Bader, 2007: Tropical sea surface temperature, vertical wind shear, and hurricane development. *Geophys. Res. Lett.,* **34,** L01710.

Murakami, H., R. Mizuta, and E. Shindo, 2012: Future changes in tropical cyclone activity projected by multi-physics and multi-SST ensemble experiments using the 60-km-mesh MRI-AGCM. *Clim. Dyn.,* **39(9-10),** 2569–2584.

Murakami, H., and coauthors, 2015: Simulation and prediction of Category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *J. Climate*, in press.

Nakano, M., M. Sawada, T. Nasuno, and M. Satoh, 2015: Intraseasonal variability and tropical cyclogenesis in the western North Pacific simulated by a global nonhydrostatic atmospheric model. *Geophys. Res. Lett.,* **42,** 565–571.

Ramsay, H. A., and A.H. Sobel, 2011: Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model. *J. Climate*, **24,** 183–193.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, and D. P. Rowell, 2003: Global analysis of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res*., **108**, 4407.

Saunders, M. A., and A. S. Lea, 2005: Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature,* **434,** 1005–1008.

538 Smith, A.B., and R. W. Katz, 2013: US billion-dollar weather and climate disasters: data
539 sources, trends, accuracy and biases. *Nat. Hazards*, **67,** 387–410.
540 Swanson, K. L., 2008: Nonlocality of Atlantic tropical cyclone intensities. *Geochem. Geophys.*
541 *Geosyst.* **9,** Q04V01.
542 Vecchi, G. A., and G. Villarini, 2014: Next season's hurricanes. *Science,* **343(6171),** 618–619.
543 Vecchi, G. A., M. Zhao, H. Wang, G. Villarini, A. Rosati, A. Kumar, I. M. Held, and R.
544 Gudgel, 2011: Statistical-dynamical predictions of seasonal North Atlantic hurricane
545 activity.
546 Vecchi, G. A., M. and co-authors, 2011: Statistical-dynamical predictions of seasonal North
547 Atlantic hurricane activity. *Mon. Wea. Rev.*, **139,** 1070–1082.
548 Vecchi, G. A., and co-authors, 2013: Multiyear predictions of North Atlantic hurricane
549 frequency: Promise and limitations. *J. Climate*, **26,** 7994–8016.
550 Vecchi, G. A., and co-authors, 2014: On the seasonal forecasting of regional tropical cyclone
551 activity. *J. Climate*, **27,** 7994–8016.
552 Villarini, G., G. A. Vecchi, and J. A. Smith, 2010: Modeling of the dependence of tropical
553 storm counts in the North Atlantic basin on climate indices. *Mon. Wea. Rev.,* **138(7),**
554 2681–2705.
555 Villarini, G., G. A. Vecchi, T. R. Knutson, M. Zhao, and J. A. Smith, 2011: North Atlantic
556 tropical storm frequency response to anthropogenic forcing: Projections and sources of
557 uncertainty. *J. Climate*, **24,** 3224–3238.
558 Villarini, G., and G. A. Vecchi, 2012: North Atlantic power dissipation index (PDI) and
559 accumulated cyclone energy (ACE): Statistical modelling and sensitivity to sea surface
560 temperature. *J. Climate*, **25,** 625–637.

561 Villarini, G., G. A. Vecchi, and J. A. Smith, 2012: U.S. landfalling and North Atlantic
562 hurricanes: Statistical modelling of their frequencies and ratios. *Mon. Wea. Rev.*, **140,** 44–
563 65.
564 Villarini, G., and G. A. Vecchi, 2013: Multiseason lead forecast of the North Atlantic power
565 dissipation index (PDI) and accumulated cyclone energy (ACE). *J. Climate*, **26,** 3631–
566 3643.
567 Vitart, F., 2006: Seasonal forecasting of tropical storm frequency using a multi-model
568 ensemble. *Quart. J. Roy. Meteor. Soc.* **132,** 647–666.
569 Vitart, F., and co-authors, 2007: Dynamically-based seasonal forecasts of Atlantic tropical
570 storm activity issued in June by EUROSIP. *Geophys. Res. Lett.,* **34,** L16815.
571 Vitart, F., 2009: Impact of the Madden Julian Oscillation on tropical cyclones and risk of
572 landfall in the ECMWF forecasting system. *Geophys. Res. Lett.,* **36,** L15802.
573 Wang, H., J.-K. E. Schemm, A. Kumar, W. Wang, L. Long, M. Chelliah, G.D. Bell, and P.
574 Peng, 2009: A Statistical Forecast Model for Atlantic Seasonal Hurricane Activity Based
575 on the NCEP Dynamical Seasonal Forecast. *J. Climate*, **22**, 4481–4500. doi:
576 http://dx.doi.org/10.1175/2009JCLI2753.1
577 Wittenberg, A. T., A. Rosati, N. –C. Lau, and J. J. Ploshay, 2006: GFDL's CM2 global
578 coupled climate models. Part III: Tropical pacific climate and ENSO. *J. Climate*, **19,** 698–
579 722.
580 Xiang, B., S.-J. Lin, M. Zhao, S. Zhang, G. Vecchi, T. Li, X. Jiang, L. Harris, J.-H. Chen,
581 2015a: Beyond weather time scale prediction for Hurricane Sandy and super typhoon
582 Haiyan in a global climate model. *Mon. Wea. Rev.*, **143,** 524–535.

583 Xiang, B., M. Zhao, X. Jiang, S.-J. Lin, T. Li, X. Fu, G. Vecchi, A. Sobel, I.-S. Kang, P.-C.
584 Hsu, E. Maloney, D. Kim, 2015b: 3-4 week MJO prediction skill in a GFDL coupled
585 model. *J. Climate,* In press.
586 Zhang, S. and A. Rosati, 2010: An inflated ensemble filter for ocean data assimilation with a
587 biased coupled GCM. *Mon. Wea. Rev.*, **138,** 3905–3931.
588 Zhao, M., I. M. Held, and G. A. Vecchi, 2010: Forecasts of the hurricane season using a global
589 atmospheric model assuming persistence of SST anomalies. *Mon. Wea. Rev.*, **138,** 3858–
590 3868.
591

592 List of Figures

593 **FIG. 1** TC tracks during the period 1980–2014 as separated by the cluster analysis. The fuzzy
594 c-mean clustering is first applied to the observed TC tracks (green), yielding mean track (red)
595 for each cluster. All predicted TC tracks by FLOR (black) are assigned to observed TC cluster
596 regardless of any lead-months and ensemble members based on root-mean-square error
597 between the predicted and mean TC track. The number in the bottom-right corner indicates the
598 sample size with the fractional ratio in parenthesis. Blue domain in each panel shows the
599 region of the United States considered for the landfall.
600

601 **FIG. 2** Forecast skill in predicting TC frequency for each cluster and for each lead month
602 predicted by the dynamical model. (a) Rank correlation between observed and predicted. (b)
603 Root-mean-square error. Black line shows the basin-total TC frequency as defined as total TC
604 frequency among the TC clusters.
605

606 **FIG. 3** Correlation map between the time-series of observed TC frequency for each cluster
607 and simulated mean large-scale parameters during July–November for each 1° × 1° grid box.
608 (a–d) Relative SST (*RSST*) defined as the local SST anomaly subtracted from the tropical
609 mean (30°S–30°N) SST anomaly. (e–h) Geopotential height at the 500 hPa level (Z_{500}); (i–l)
610 Zonal component of vertical wind shear (200–850 hPa, *WS*). Rectangles indicate domains for
611 predictors as red rectangles showing positive sign and blue rectangles showing negative sign.
612
613 **FIG. 4** Correlation map between *RSST* and CL3 TC frequency during July–November. (a)
614 Correlation between observed *RSST* and predicted TC frequency by the model for CL3 at lead-
615 month 0. (b) As in (a), but for correlation between observed *RSST* and observed TC frequency
616 for CL3. Red (blue) rectangles indicate domains for predictors showing a positive (negative)
617 sign.
618
619 **FIG. 5** Anomaly correlation for simulated large-scale parameters for each lead-month of L=0,
620 3, and 6. (a–c) *RSST*, (d–f) Z_{500}, and (g–i) *WS*. The rectangles in these domains are the same as
621 in Fig. 3.
622
623 **FIG. 6** Results of interannual variation of TC frequency by (a–d) the regression and (e–h) the
624 leave-one-out cross validation (LOOCV) for each cluster at lead-month 0. Observed
625 (Regressed or cross validated) TC frequency is shown in black (blue). Blue regions indicates
626 range of 10% bottom range and 90% top range computed from random resampling based on
627 the Poisson distribution. Numbers shown for each panel show rank correlation and RMSE
628 between the black and blue lines. The star mark indicates statistical significance at 99% level.

FIG. 7 (a, b) Comparisons of forecast kills between dynamical model (solid line) and hybrid model (dashed line) in predicting TC frequency for each cluster and for each lead month. (a) shows the rank correlation between observations and models, whereas (b) shows RMSE between them. (c, d) As in (a, b), but for basin-total TC frequency.

FIG. 8 Observed time-series of TC frequency and landfall ratio in the North Atlantic. The red line indicates observed basin-total TC frequency, whereas the blue line indicates observed landfall TCs over the United States. The black line showss the landfall ratio. The green line shows Niño-3.4 index which is obtained from the mean SST anomaly in the region bounded by 5°N and 5°S, and between 170°W to 120°W.

FIG. 9 Correlation map between observed TC landfall ratio over the United States and simulated SST anomalies for each lead month. (a) Lead month 0, (b) Lead month 3, and (c) Lead month 6.

FIG. 10 Results of interannual variation of TC landfall ratio by (a–c) the regression and (d–f) the leave-one-out cross validation (LOOCV) for each lead month of (a, d) 0, (b, e) 3, and (c, f) 6. Blue areas indicate the range between the 10% and 90% computed from random resampling based on the binomial distribution. Numbers shown in the middle top for each panel indicate rank correlation and RMSE between the black and blue lines. The star mark indicates statistical significance at 99% level.

652 **FIG. 11** As in Fig. 7c,d, but for TC landfall ratio over the United States.

653

654 **FIG. 12** Schematic diagram showing synthesized hybrid model to predict landfall TC
655 frequency. Details are explained in the main text.

656

657 **FIG. 13** Results of the interannual variation in the frequency of landfall TCs by (a–c) the
658 dynamical model and (d–f) the synthesized hybrid model for each lead month of (a, d) 0, (b, e)
659 2, and (c, f) 4. Blue areas indicate range of 10% bottom range and 90% top range computed
660 from random resampling. Numbers shown for each panel show rank correlation and RMSE
661 between the black and blue lines. The star mark indicates statistical significance at 99% level.

662

663 **FIG. 14** As in Fig. 7c,d, but for frequency of landfall TCs over the United States.

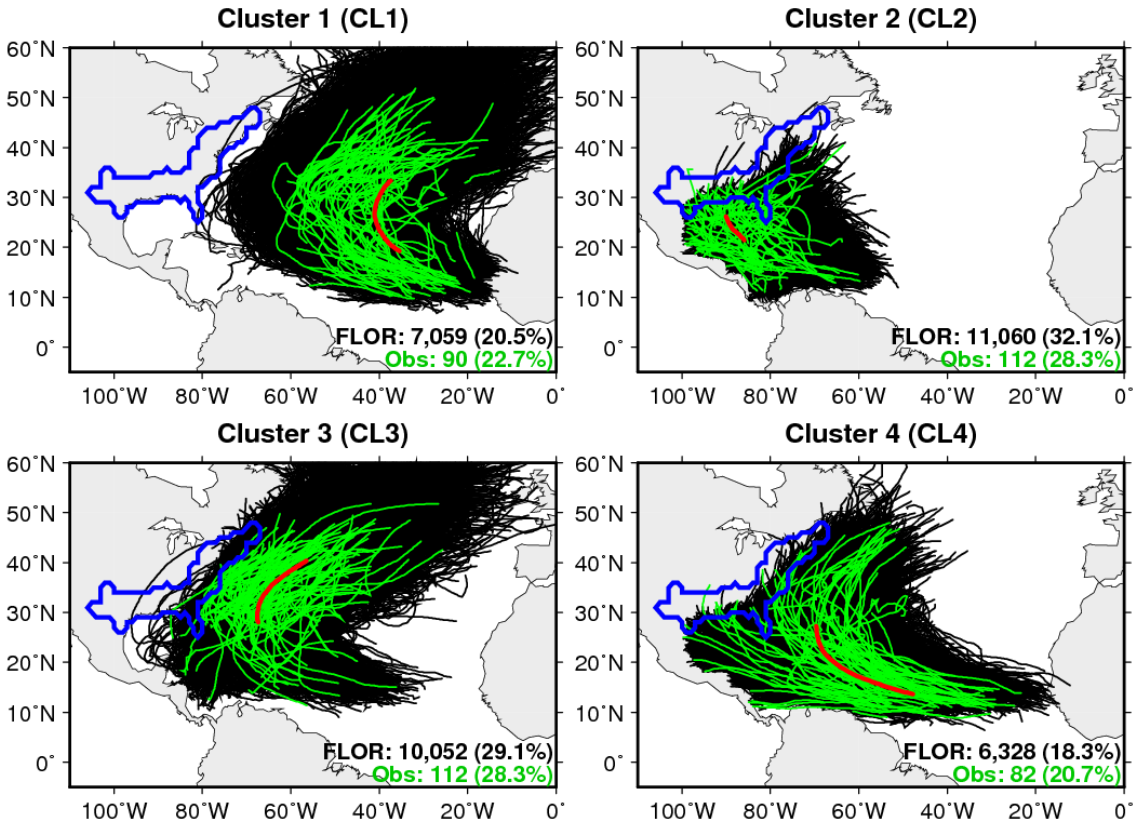

FIG. 1 TC tracks during the period 1980–2014 as separated by the cluster analysis. The fuzzy c-mean clustering is first applied to the observed TC tracks (green), yielding mean track (red) for each cluster. All predicted TC tracks by FLOR (black) are assigned to observed TC cluster regardless of any lead-months and ensemble members based on root-mean-square error between the predicted and mean TC track. The number in the bottom-right corner indicates the sample size with the fractional ratio in parenthesis. Blue domain in each panel shows the region of the United States considered for the landfall.

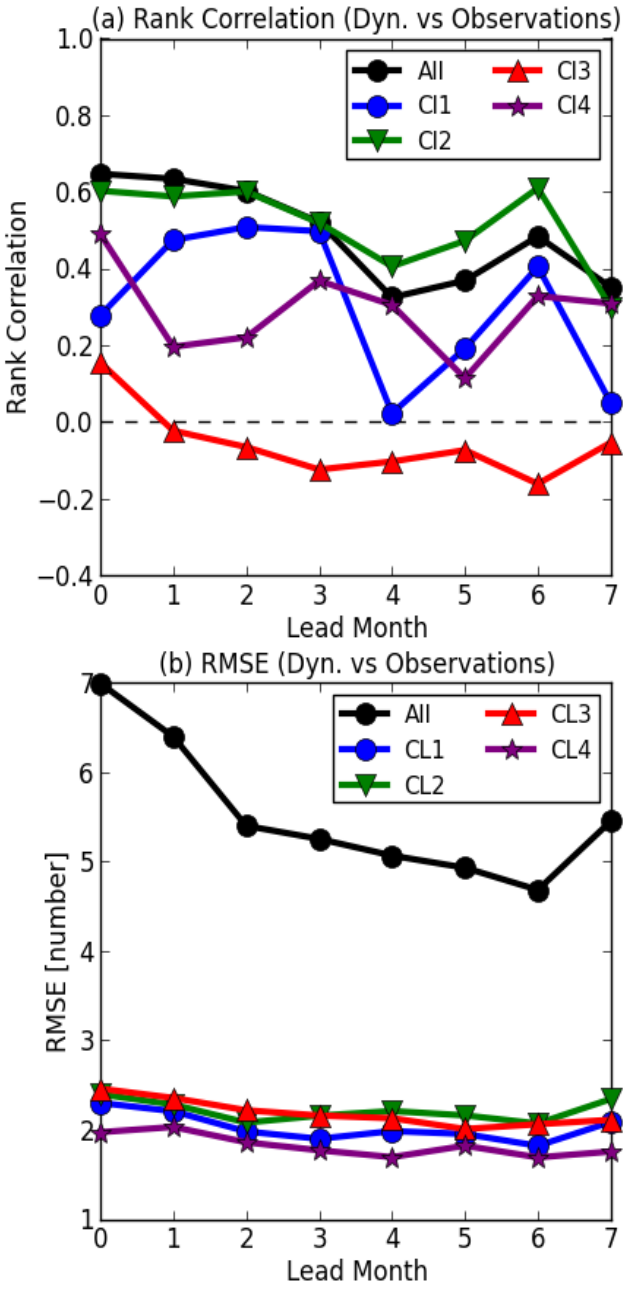

FIG. 2 Forecast skill in predicting TC frequency for each cluster and for each lead month predicted by the dynamical model. (a) Rank correlation between observed and predicted. (b) Root-mean-square error. Black line shows the basin-total TC frequency as defined as total TC frequency among the TC clusters.

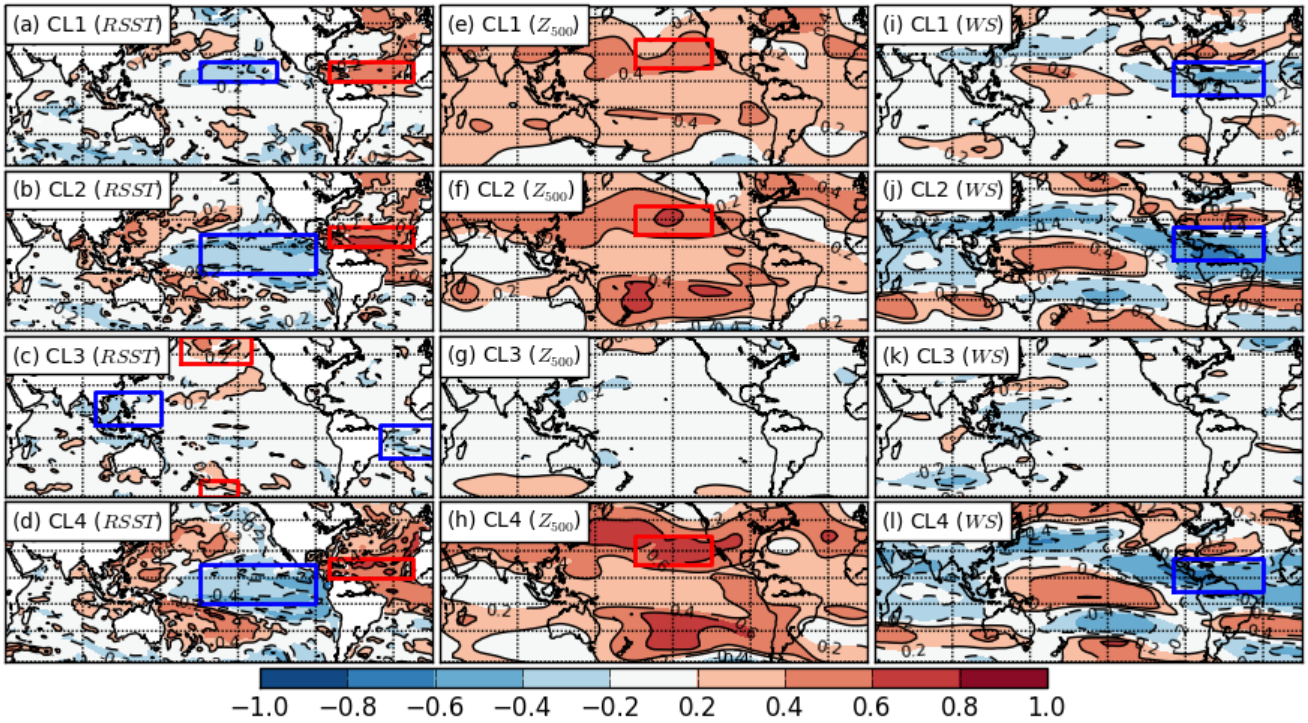

FIG. 3 Correlation map between the time-series of observed TC frequency for each cluster and simulated mean large-scale parameters during July–November for each 1° × 1° grid box. (a–d) Relative SST (*RSST*) defined as the local SST anomaly subtracted from the tropical mean (30°S–30°N) SST anomaly. (e–h) Geopotential height at the 500 hPa level (Z_{500}); (i–l) Zonal component of vertical wind shear (200–850 hPa, *WS*). Rectangles indicate domains for predictors as red rectangles showing positive sign and blue rectangles showing negative sign.

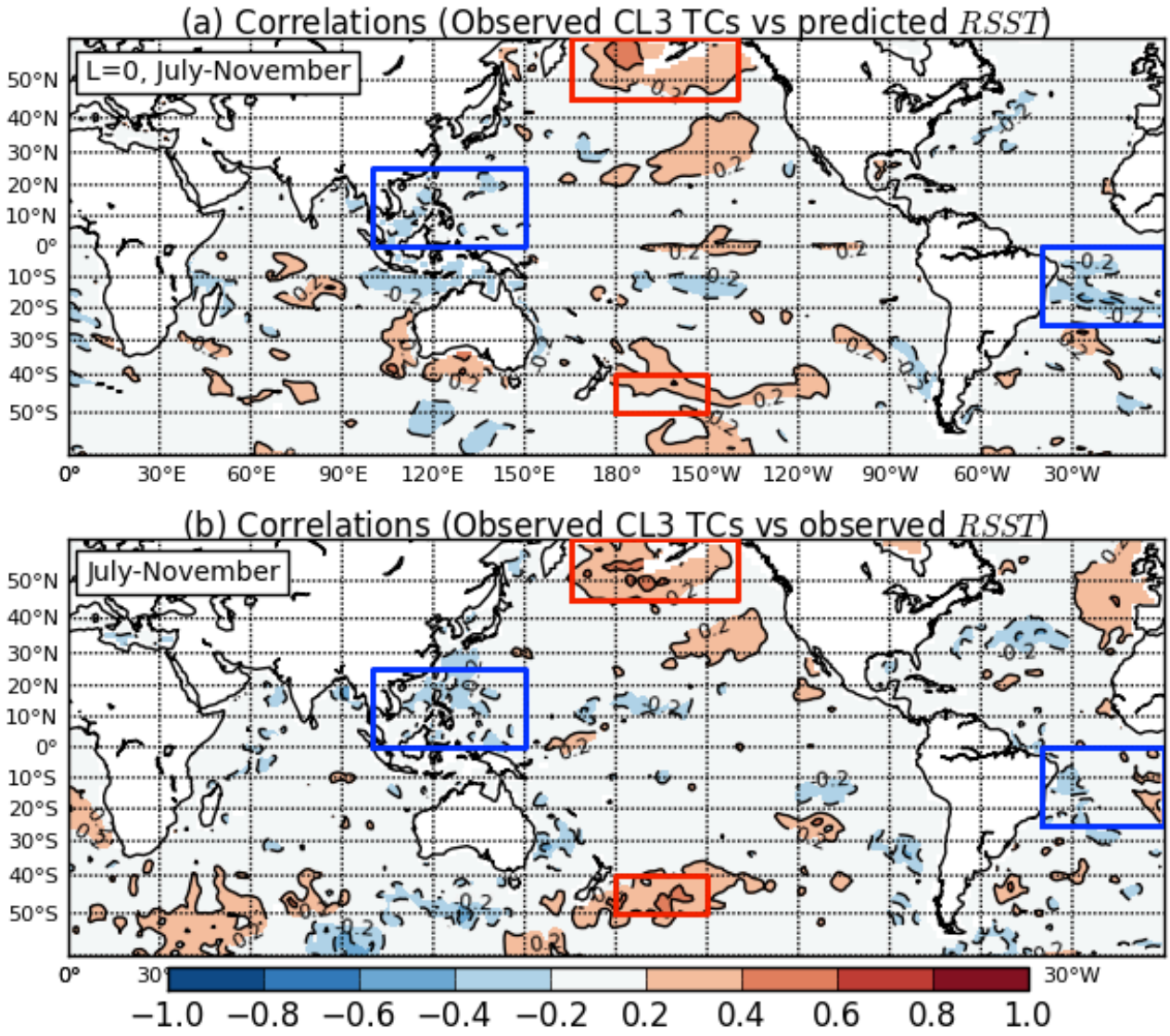

FIG. 4 Correlation map between *RSST* and CL3 TC frequency during July–November. (a) Correlation between observed *RSST* and predicted TC frequency by the model for CL3 at lead-month 0. (b) As in (a), but for correlation between observed *RSST* and observed TC frequency for CL3. Red (blue) rectangles indicate domains for predictors showing a positive (negative) sign.

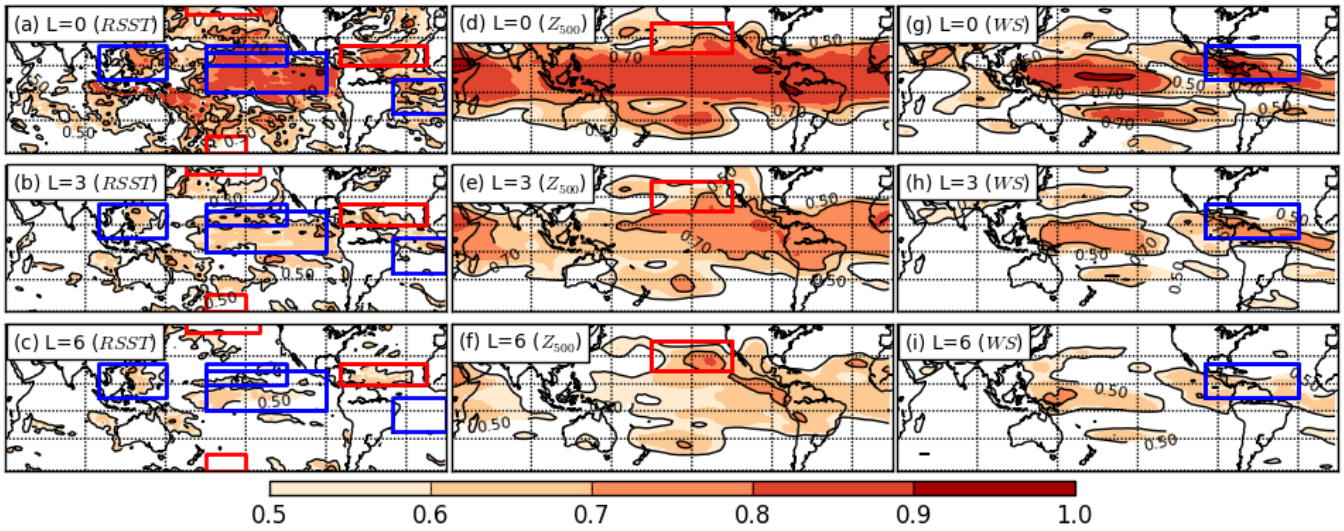

FIG. 5 Anomaly correlation for simulated large-scale parameters for each lead-month of L=0, 3, and 6. (a–c) *RSST*, (d–f) Z_{500}, and (g–i) *WS*. The rectangles in these panels are the same as in Fig. 3.

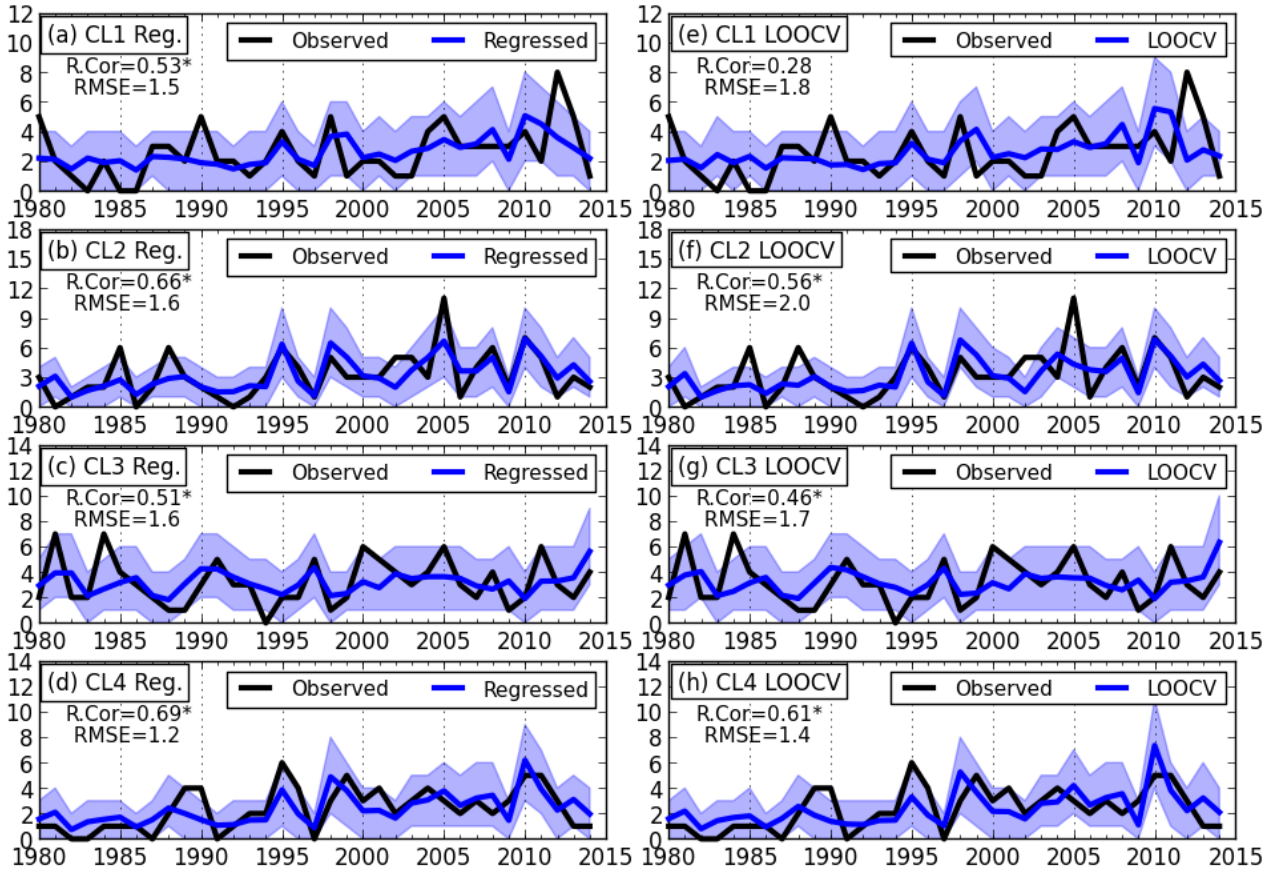

FIG. 6 Results of interannual variation of TC frequency by (a–d) the regression and (e–h) the leave-one-out cross validation (LOOCV) for each cluster at lead-month 0. Observed (Regressed or cross validated) TC frequency is shown in black (blue). Blue regions indicates range of 10% bottom range and 90% top range computed from random resampling based on the Poisson distribution. Numbers shown for each panel show rank correlation and RMSE between the black and blue lines. The star mark indicates statistical significance at 99% level.

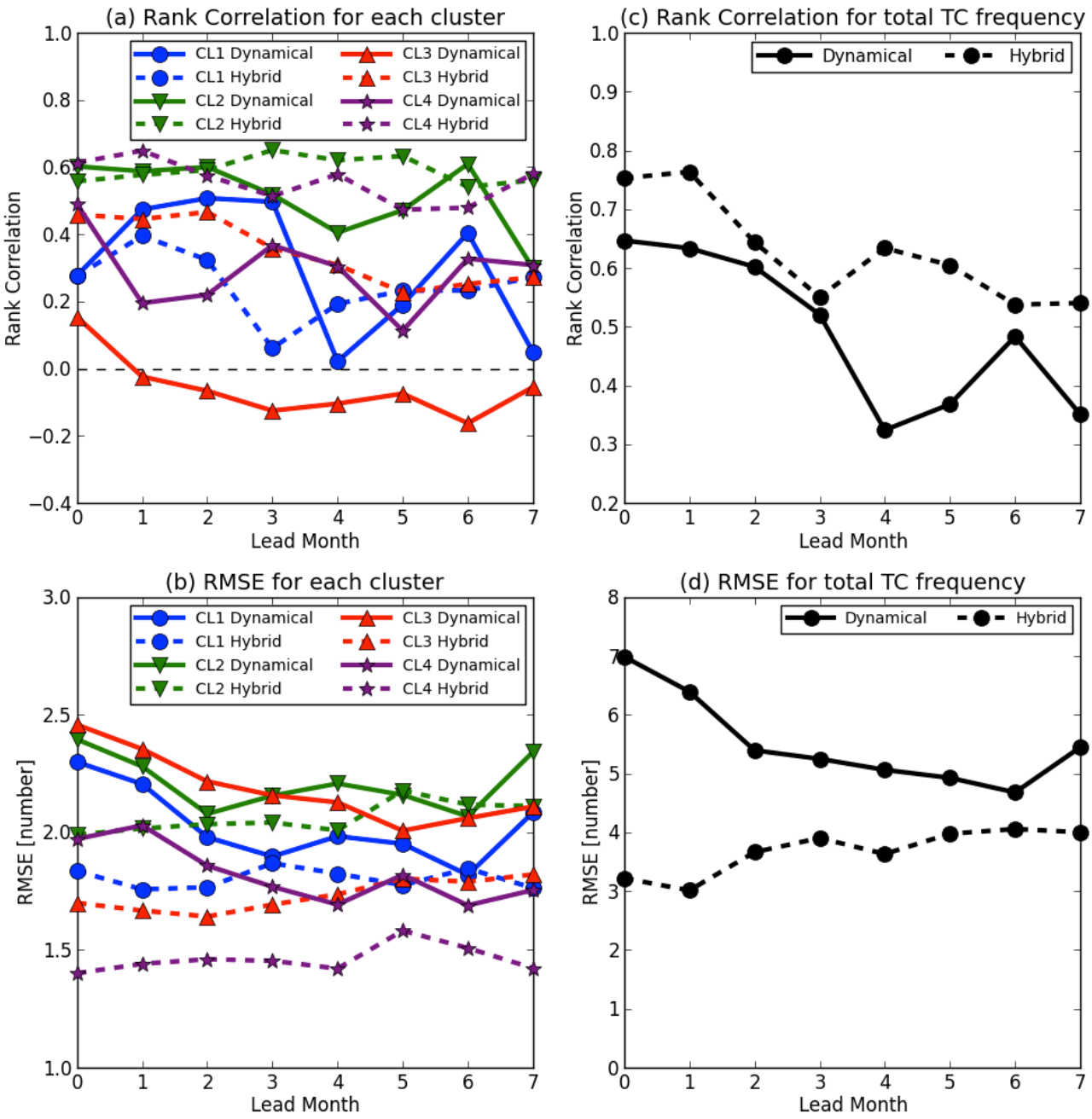

FIG. 7 (a, b) Comparisons of forecast kills between dynamical model (solid line) and hybrid model (dashed line) in predicting TC frequency for each cluster and for each lead month. (a) shows the rank correlation between observations and models, whereas (b) shows RMSE between them. (c, d) As in (a, b), but for basin-total TC frequency.

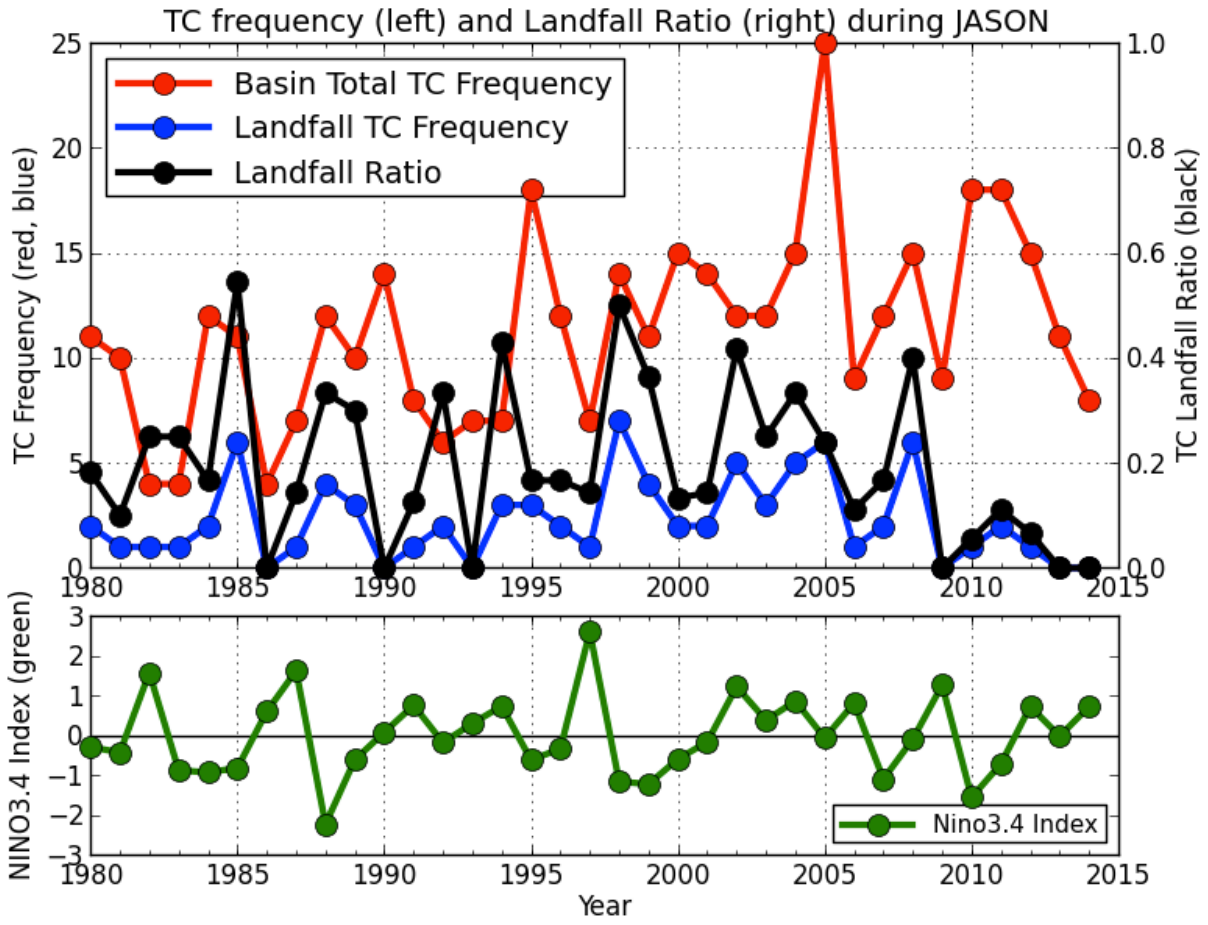

FIG. 8 Observed time-series of TC frequency and landfall ratio in the North Atlantic. The red line indicates observed basin-total TC frequency, whereas the blue line indicates observed landfall TCs over the United States. The black line shows the landfall ratio. The green line shows Niño-3.4 index which is obtained from the mean SST anomaly in the region bounded by 5°N and 5°S, and between 170°W to 120°W.

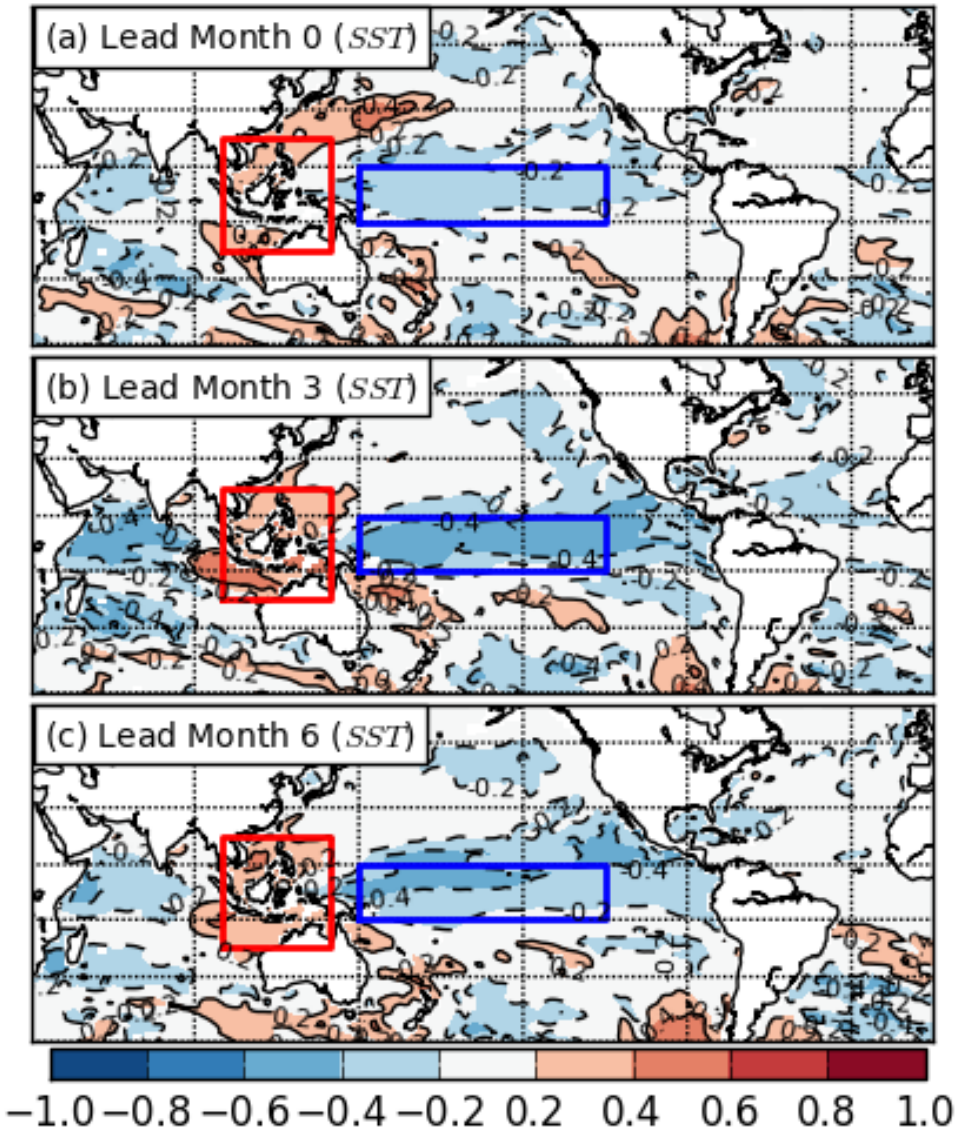

FIG. 9 Correlation map between observed TC landfall ratio over the United States and simulated SST anomalies for each lead month. (a) Lead month 0, (b) Lead month 3, and (c) Lead month 6.

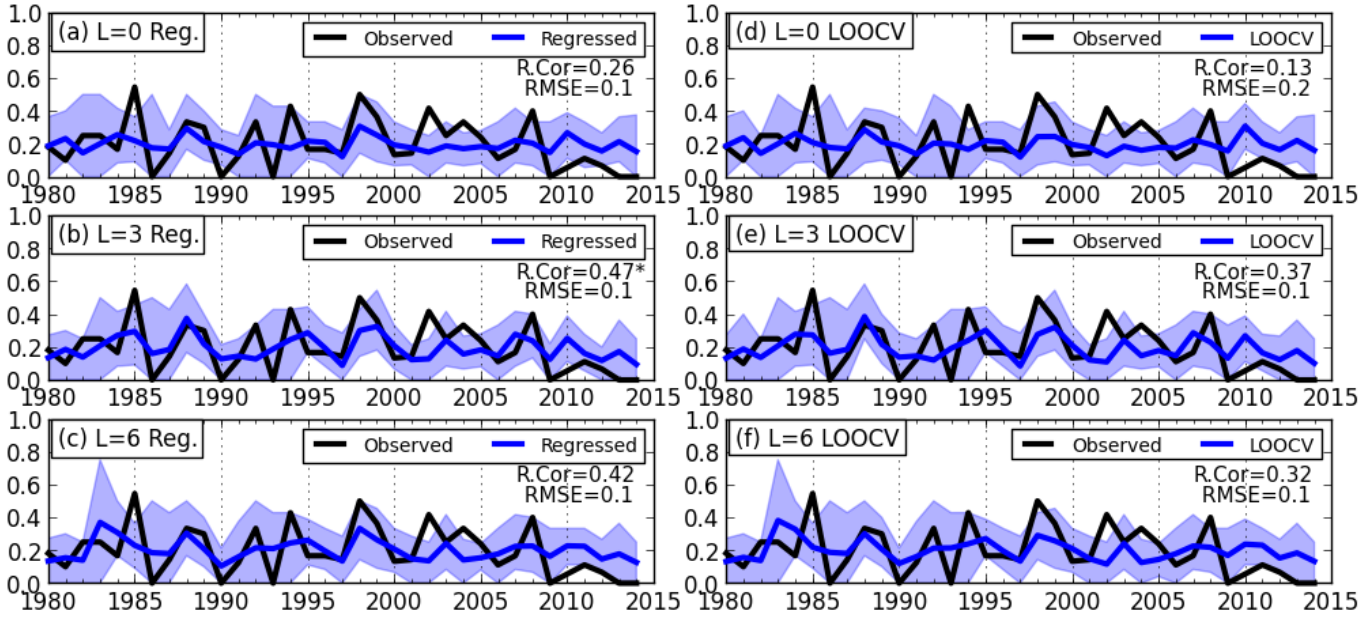

FIG. 10 Results of interannual variation of TC landfall ratio by (a–c) the regression and (d–f) the leave-one-out cross validation (LOOCV) for each lead month of (a, d) 0, (b, e) 3, and (c, f) 6. Blue areas indicate the range between the 10% and 90% computed from random resampling based on the binomial distribution. Numbers shown in the middle top for each panel indicate rank correlation and RMSE between the black and blue lines. The star mark indicates statistical significance at 99% level.

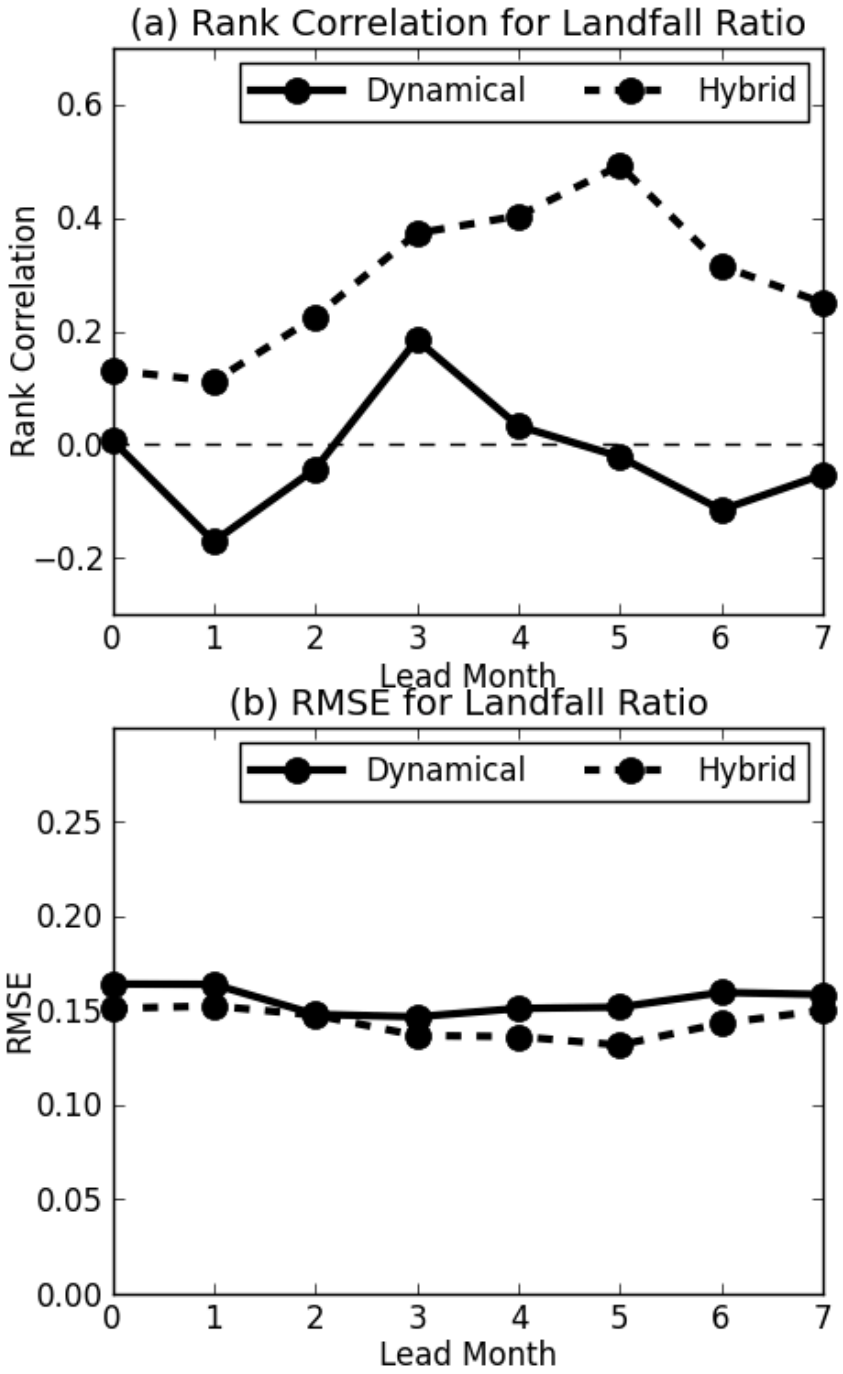

FIG. 11 As in Fig. 7c,d, but for TC landfall ratio over the United States.

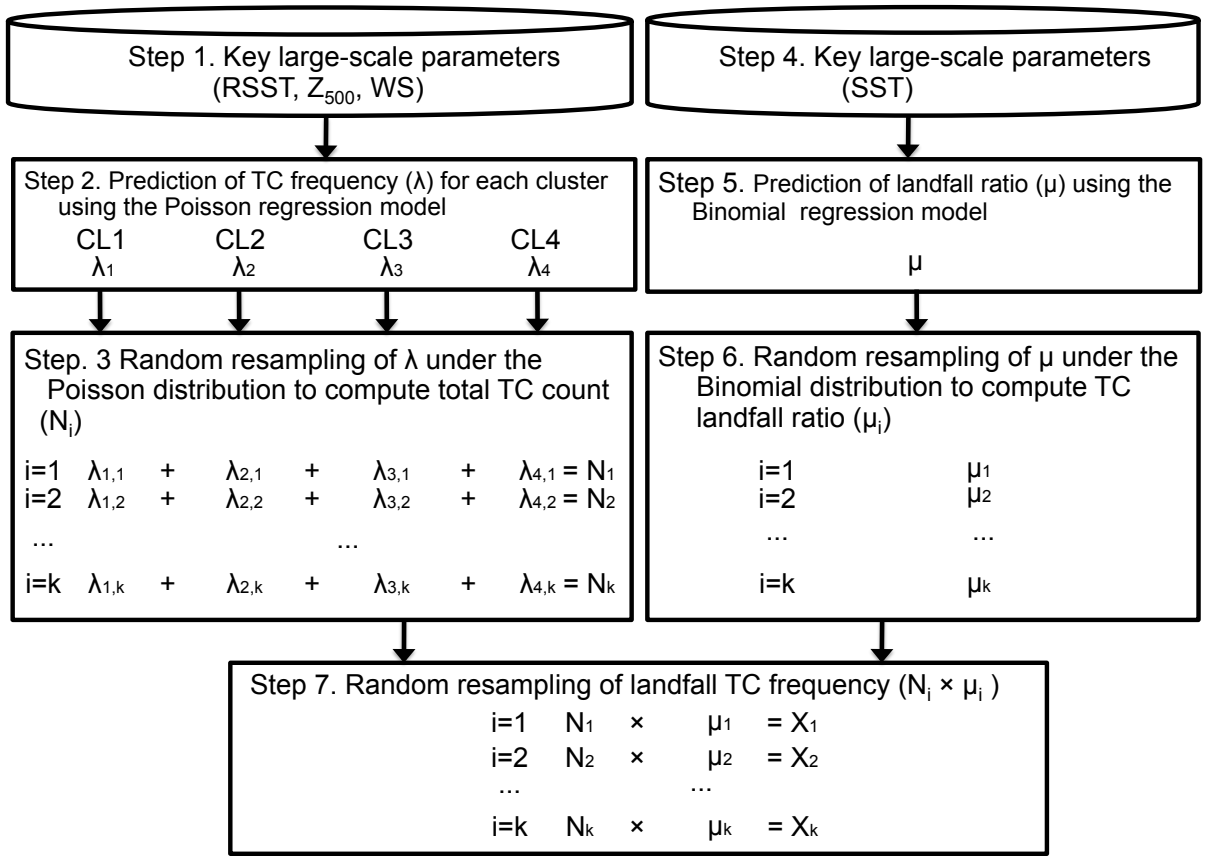

FIG. 12 Schematic diagram showing synthesized hybrid model to predict landfall TC frequency. Details are explained in the main text.

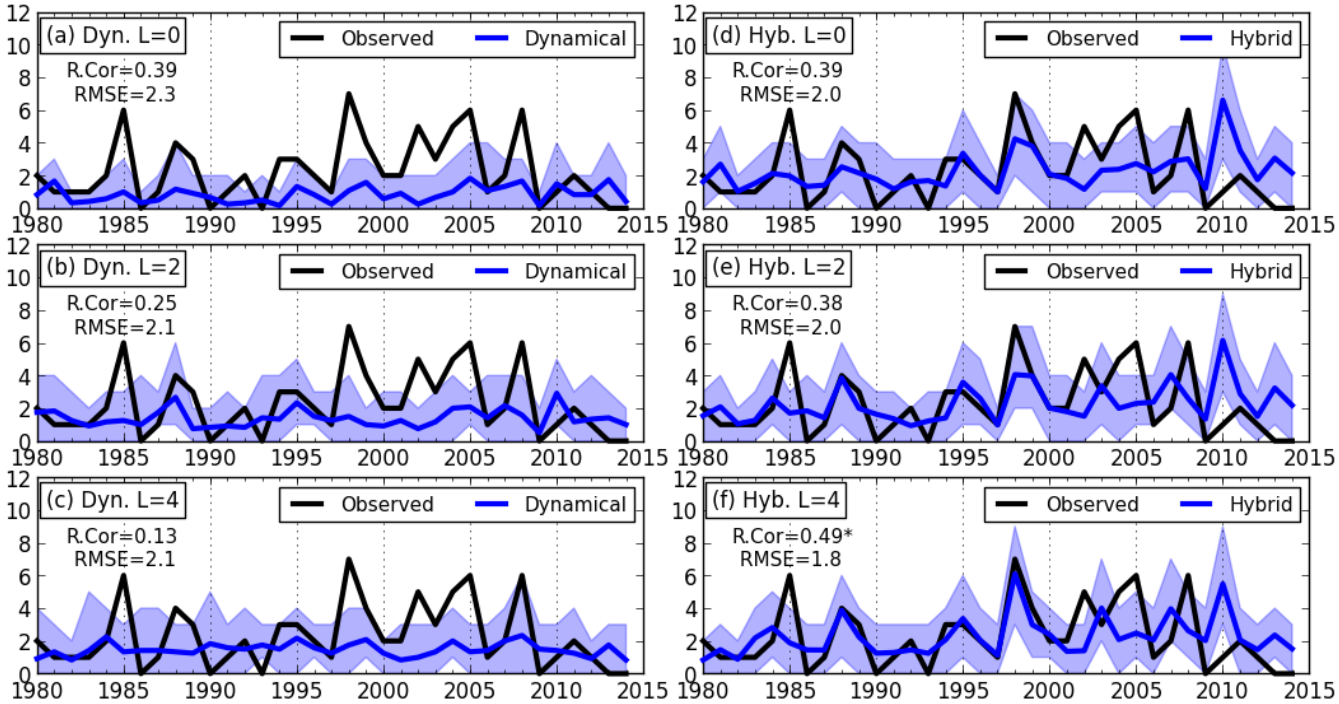

FIG. 13 Results of the interannual variation in the frequency of landfall TCs by (a–c) the dynamical model and (d–f) the synthesized hybrid model for each lead month of (a, d) 0, (b, e) 2, and (c, f) 4. Blue areas indicate range of 10% bottom range and 90% top range computed from random resampling. Numbers shown for each panel show rank correlation and RMSE between the black and blue lines. The star mark indicates statistical significance at 99% level.

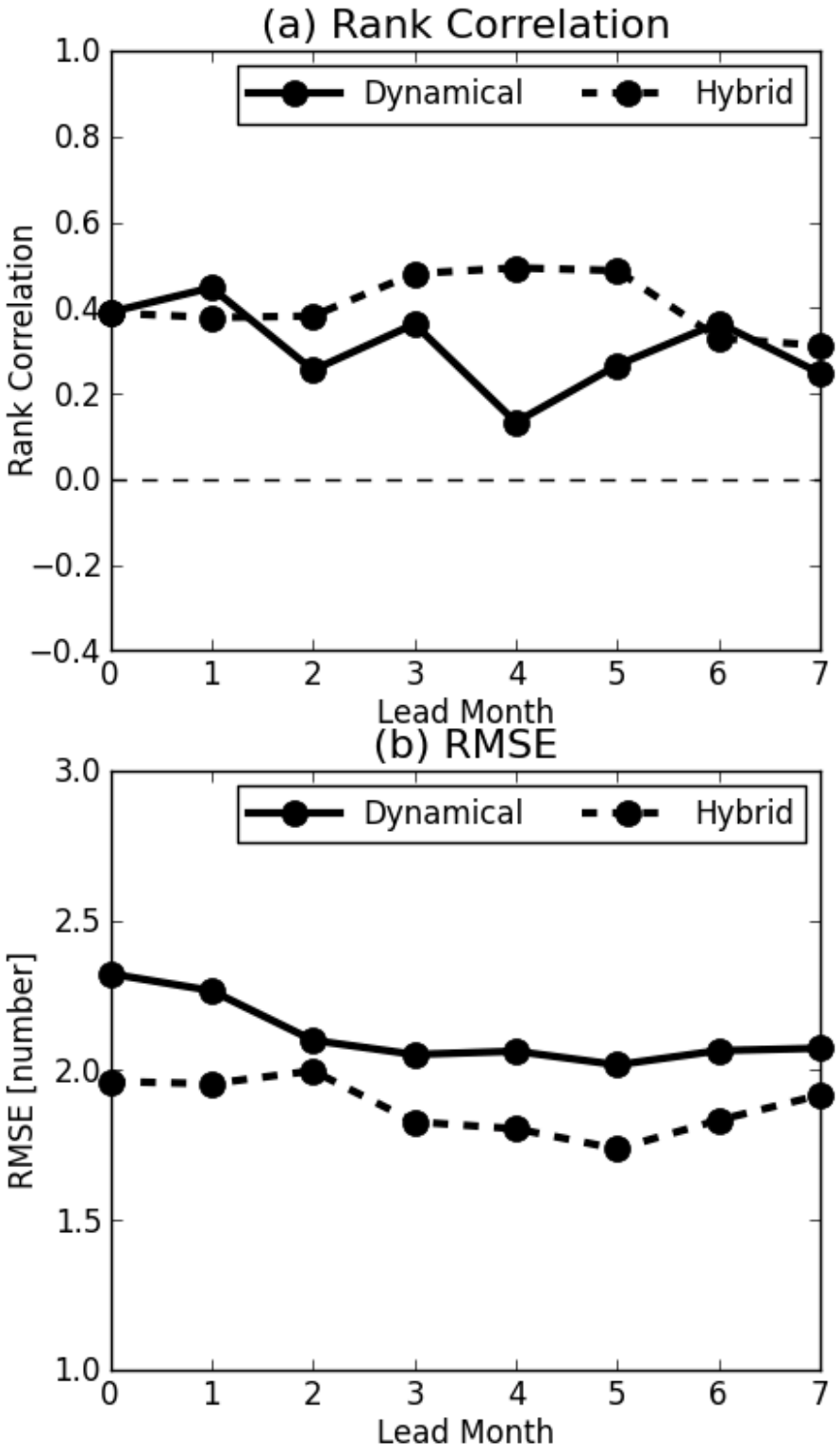

FIG. 14 As in Fig. 7c,d, but for frequency of landfall TCs over the United States.